

A Comparative Study of Covariance and Precision Matrix Estimators for Portfolio Selection

M. Senneret¹, Y. Malevergne^{2,3}, P. Abry⁴, G. Perrin¹, L. Jaffrès¹

¹ Vivienne Investissement, Lyon, France

² Coactis EA 4161 – Université de Lyon - Université de Saint-Etienne, France

³ EMLYON Business School, France

⁴ CNRS, ENS Lyon, France

Preliminary draft. Do not quote

Abstract

We conduct an empirical analysis of the relative performance of several estimation methods for the covariance and the precision matrix of a large set of European stock returns with application to portfolio selection in the mean-variance framework. We develop several precision matrix estimators and compare their performance to their covariance matrix estimators counterpart. We account for the presence of short-sale restrictions, or the lack thereof, on the optimization process and study their impact on the stability of the optimal portfolios. We show that the best performing estimation strategy, on the basis of the *ex-post* Sharpe ratio, does not actually depend on the fact that we choose to estimate the covariance or the precision matrix. Nonetheless, the optimal portfolios derived from the estimated precision matrix enjoy a much lower turnover rate and concentration level even in the absence of constraints on the investment process.

Keywords: Portfolio selection, covariance matrix, precision matrix, multivariate estimation, shrinkage, sparsity.

JEL Code: C13, C51, C61, G11, G15.

1 Introduction

The estimation of the covariance matrix of assets returns is an important step for a successful implementation of the mean-variance portfolio optimization approach (Elton and Gruber 1973, DeMiguel *et al.* 2009). However the estimation of large covariance matrices is a notoriously difficult task. Actually, if the pointwise convergence of the usual estimators is guaranteed under mild assumptions met in real-life conditions, their eigenvalues and eigenvectors often remain quite noisy, testifying of a significant loss of information during the estimation process. Besides, the inversion of large matrices is a tedious task, numerically unstable when the matrix is ill-conditioned as is usually the case in practice when the number of observations is (in the most favorable case) close to the number of assets under consideration. Hence matrix inversion may contribute to add some noise during the optimization step and to make the mean-variance approach useless due to its tendency to maximize the effects of errors in the input assumptions (Michaud 1989). In fact, the estimation error and the numerical instability are so large that DeMiguel *et al.* (2009) concluded that the naive equally-weighted investment scheme is superior to the optimal asset allocation derived from the mean-variance method both in terms of Sharpe ratio and certainty equivalent. For the US stock market, based on monthly returns, they estimated that the sample size needed for the mean-variance strategy to outperform the equally-weighted portfolio is larger than 3,000 months for a portfolio with 25 assets and 6,000 months for a portfolio made of 50 assets. This should lead one to pessimistically conclude with these authors that “*there are still many miles to go before the gains promised by optimal portfolio choice can actually be realized out-of-sample.*”

However, Ledoit and Wolf (2004b) report significantly lower *ex-post* variances for global minimum-variance portfolios (GMVP) derived from shrunked sample covariance matrices compared with the *ex-post* variance of the equally-weighted portfolio or the GMVP obtained on the basis of the raw sample covariance matrix. These findings are confirmed by Disatnik and Benninga (2007) which also suggest that the way the sample covariance matrix is shrunked is not so important as far as the *ex-post* variance only is concerned. In addition to the *ex-post* variance of the GMVP obtained by shrinkage methods, Jagannathan and Ma (2003) focus on their out-of-sample performance in terms of Sharpe ratio. Their results contrast with those obtained by DeMiguel *et al.* (2009) in so far as, in the absence of investment restrictions, the GMVP derived from the shrinkage methods exhibit *ex-post* Sharpe ratios larger than the one of the equally-weighted strategy. Hence, any hope for achievable gains from the optimal portfolio theory is maybe not out of reach.

It is nonetheless surprising that, at the noticeable exception of the recent paper by Kourtis *et al.* (2012), all the efforts developed during the last decade to provide a better assessment of

the optimal portfolios in the mean-variance framework has been devoted to the improvement of the covariance matrix estimation while, in this context, the very input parameter of interest is not the covariance matrix itself but its inverse, i.e., the so-called precision matrix. Actually, the inverse of the estimated covariance matrix is expected to provide a rather poor estimate of the precision matrix. Not only because of the numerical instability of the inversion process, but also because it is well-known that the inverse of the unbiased sample covariance matrix only provides a (severely) biased estimator on the inverse covariance matrix (Muirhead 2005). Hence, for a portfolio made of 50 assets and a sample size of 100 observations, the precision matrix estimated by direct inversion of the sample covariance matrix is, on average, twice as large as the population precision matrix. In this respect, it is important to develop reliable estimators of the precision matrix and to compare their accuracy, in terms of portfolio performance, with the results obtained from the inversion of the estimated covariance matrix. It is the primary goal of this paper.

In addition, it seems that the introduction of restrictions on short-sales improves the performance of the GMVP, lowering the out-of-sample risk. It is important to understand to what extent this result is related to the instability of the inversion of the estimated covariance matrix. Indeed, Jagannathan and Ma (2003) show that the introduction of investment restrictions acts as a shrinkage of the covariance matrix and thus leads to deal with an “effective” covariance matrix whose conditioning is better and, hence, enjoys better numerical properties and exhibits a more suitable behavior during the optimization process. In this respect, Jagannathan and Ma (2003) suggest that the introduction of wrong constraints on the investment strategy may actually help improve the portfolio performance. Pushing this argument one step further, we can wonder whether the estimation of the precision matrix still makes the addition of unnecessary constraints relevant or if it is enough, in and of itself, to stabilize the optimization. It is the secondary goal of the paper.

The paper is organized as follows. In the next section, we recall the basic framework for the mean-variance approach with and without restrictions on short-sales and we fix the main notations. In section 3, we present the different estimation strategies considered here. We briefly survey relevant results for covariance matrix estimation and state new results relative to the estimation of the precision matrix. In section 4, we report the results of our empirical analysis conducted on a large dataset of stocks comprised in the Euro Stoxx 600 index over the last decade. We briefly conclude in section 5. All proofs and technical details are gathered in appendix.

2 Minimum variance portfolio allocation

Given a set of p risky securities, with a full-rank $p \times p$ covariance matrix of assets returns denoted by Σ , the global minimum variance portfolio without restriction on short-sales is solution to the optimization program

$$\begin{aligned} w^* &= \operatorname{argmin}_w w' \Sigma w \\ &\text{s.t. } 1_p' w = 1 \end{aligned} \quad (1)$$

where 1_p is the p -vector of ones and \cdot' denotes the transpose operator.

The solution to the problem (1) is well-known

$$w^* = \frac{\Theta 1_p}{1_p' \Theta 1_p}, \quad (2)$$

where $\Theta = \Sigma^{-1}$ denotes the $p \times p$ precision matrix of the assets returns. The only practical issue we have to care about is the way to reliably estimate the weights of this portfolio. Indeed shall we prefer the estimator

$$\hat{w}^* = \frac{\hat{\Theta} 1_p}{1_p' \hat{\Theta} 1_p} \quad \text{or} \quad \hat{w}^* = \frac{\hat{\Sigma}^{-1} 1_p}{1_p' \hat{\Sigma}^{-1} 1_p} ? \quad (3)$$

Intuition should suggest to prefer the first one for accuracy reasons, but common practice promotes the use of the second one. An obvious motivation is its ease of implementation. It is therefore important to investigate their relative properties for different estimation strategies.

Now, when short-sales are restricted, and even forbidden, the optimization program reads

$$\begin{aligned} w^* &= \operatorname{argmin}_w w' \Sigma w \\ &\text{s.t. } 1_p' w = 1 \\ &\quad w \geq 0 \end{aligned} \quad (4)$$

where the inequality is to be taken as a generalized inequality over the nonnegative orthant. Jagannathan and Ma (2003) showed that the constraints can make the optimization problem more robust and explain why introducing wrong constraints may help reduce the global portfolio risk. Actually, the introduction of constraints acts as the shrinkage of the covariance matrix, which allows rationalizing the observation reported by Pantaleo *et al.* (2011), among others, that the choice of the covariance matrix estimator is not very important in the presence of short-sale restrictions. To understand this point, let us consider the Lagrangian of the constrained problem (4)

$$L(w, \lambda, \nu) = w' \Sigma w + \lambda (1_p' w - 1) - \nu' w, \quad (5)$$

where $\lambda \in \mathbb{R}$ and $\nu \in \mathbb{R}_+^p$ are the Lagrange multipliers associated to the two sets of constraints. Writing down the Karush-Kuhn-Tucker conditions of optimality, we get

$$2\Sigma w + \lambda \cdot 1_p - \nu = 0, \quad (6a)$$

$$1'w - 1 = 0, \quad (6b)$$

$$\text{Diag}(\nu) w = 0, \quad (6c)$$

$$w, \nu \geq 0. \quad (6d)$$

Hence equation (6a) together with equations (6b-6c) is equivalent to

$$2 \left[\Sigma + \text{Diag}(\nu) - \frac{1}{2} \cdot (1_p \nu' + \nu 1_p') \right] w + \lambda \cdot 1_p = 0, \quad (7)$$

so that, with a slightly different expression than the one mentioned by Jagannathan and Ma (2003), the optimal portfolio with short-sale restrictions is the same as the unconstrained optimal portfolio with the actual covariance matrix Σ replaced by the “effective” covariance matrix $\Sigma + \text{Diag}(\nu) - \frac{1}{2} \cdot (1_p \nu' + \nu 1_p')$. This additional term leads to a decay in the effective correlations between the assets. Indeed, the diagonal terms of the additional matrix are all zero, so that the individual variances remain unchanged while the correlation between assets i and j are shifted by the amount $-\frac{\nu_i + \nu_j}{2 \cdot \sqrt{\Sigma_{ii} \cdot \Sigma_{jj}}}$. As a consequence, the actual diversification increases and the weights of the constrained optimal portfolio are more spread out over the different assets.

As in the absence of constraints, the solution to the constrained optimization problem involves the estimation of the covariance matrix or of the precision matrix. In this respect, it is sensible to compare the relative performance of both approaches. The results of our empirical study will be reported in section 4. Before that, let us expose the different estimation methods we will compare.

3 Covariance and precision estimation strategies

In this section we present the main estimation approaches retained in the paper and just provide a brief overview of their main properties. We refer the reader to Bai and Shi (2011) for an in-depth survey of alternative estimation strategies for the covariance matrix. We also introduce new results regarding the estimation of the precision matrix. When needed, we will refer to p as the number of assets in the portfolio and to n as the sample size, i.e., the number of observations from which the covariance/precision matrix can be estimated.

3.1 Direct sample estimates

The sample covariance matrix estimator is certainly the most simple estimator one can consider. However, it suffers from two main deficiencies. First of all, when the number of observations n is less than the number of assets p , the sample covariance matrix is not full rank, hence it is not invertible. In such a case, the Moore-Penrose generalized inverse is usually retained to estimate the inverse covariance matrix. Secondly, even if the sample covariance matrix is full rank, its inverse only provides a biased estimator of the inverse population covariance matrix. Many simple, and sometimes naive but efficient, alternatives have been proposed. Among many others, let us refer to the replacement of the sample covariance matrix by a scalar matrix, which leads to the equally-weighted portfolio as the global minimum variance portfolio, or by the diagonal matrix of the sample variances or by the covariance matrix derived from a constant correlation model (Elton and Gruber 1973). The sample covariance matrix and its (generalized) inverse will provide the benchmark strategy for the horse race exposed in section 4.

Alternatively, in order (i) to reduce the noise in the sample covariance matrix, (ii) to get a full rank estimate of the covariance matrix even when the number of assets is larger than the number of observations and (iii) to get a reliable estimate of the inverse covariance matrix, the factor models can provide a simple approach. The simplest case is derived from the Sharpe (1963) market model in which the return on the market portfolio is assumed to be the single relevant factor. More general models based on the Fama and French (1992) three factors model or the Carhart (1997) four factors model can provide better approximations to the actual covariance matrix for stock returns. Alternatively, when the factors are unknown or unobservable, an approximate factor structure (Chamberlain and Rothschild 1983) can be reconstructed from a principal component analysis or a singular value analysis (Connor 1982, Bai and Ng 2002, Malevergne and Sornette 2004). For simplicity we will consider the single index market model as representative of this second class of models:

$$r_t = \alpha + \beta \cdot r_{m,t} + \varepsilon_t, \quad (8)$$

where r_t denotes the p -vector of assets returns on day $t \in \{1, \dots, n\}$, $r_{m,t}$ is the return on the market index on day t , α and β are the p -vectors of intercepts and factor loadings while ε_t is the p -vector of residuals with diagonal covariance matrix Δ . For simplicity, we assume the iid-ness of the observations. The covariance and precision matrices read

$$\Sigma = \Delta + \sigma_m^2 \beta \beta' \quad \text{and} \quad \Theta = \Delta^{-1} - \frac{\Delta^{-1} \beta \cdot \beta' \Delta^{-1}}{\frac{1}{\sigma_m^2} + \beta' \Delta^{-1} \beta}, \quad (9)$$

where $\sigma_m^2 = \text{Var } r_{m,t}$.

Based on the OLS estimate of β and the unbiased estimate of Δ , we easily get an unbiased estimator of the covariance matrix

$$\hat{\Sigma} = \frac{n-2}{n-1} \cdot \hat{\Delta} + \hat{\sigma}_m^2 \hat{\beta} \hat{\beta}', \quad (10)$$

where $\hat{\sigma}_m^2$ is the unbiased estimate of σ_m^2 . We can notice that it only differs from the naive plug-in estimator by the multiplicative factor $(n-2)/(n-1)$ which becomes quickly close to one as soon as n increases. In this respect, there is no significant difference to expect between this estimator and the pug-in estimator as soon as the number of observation is not ridiculously small. In particular, the statistical properties of the estimator are independent from the number of assets in portfolio.

We can also propose an alternative estimator to the plug-in estimator of the precision matrix (see Appendix B for the derivation):

$$\hat{\Theta} = \frac{n-4}{n-2} \hat{\Delta}^{-1} + \frac{\left(\frac{n-4}{n-2}\right)^2 \hat{\sigma}_m^2 \hat{\Delta}^{-1} \hat{\beta} \hat{\beta}' \hat{\Delta}^{-1} - \frac{1}{n-1} \hat{\Delta}^{-1}}{1 + \frac{n-4}{n-2} \hat{\sigma}_m^2 \hat{\beta}' \hat{\Delta}^{-1} \hat{\beta} - \frac{p}{n-1}}. \quad (11)$$

Here also, it is interesting to notice that, in the limit n goes to infinity, the estimator converges toward the plug-in estimator but, on the contrary to the previous case, even for large n , if the number of assets p remains large so that $p(n)/n \rightarrow \gamma > 0$, the correction in the denominator does not vanish and this estimator still significantly departs from the plug-in estimator.

To conclude this survey of simple estimators, let us mention the latent factors approach which is, to a large extend, related to the so-called random matrix theory introduced by Wigner (1953) and recently brought back to the front of the scene by Laloux *et al.* (1999, 2000) and their followers. As for the principal component analysis, the idea consists in the identification of the eigenvalues and the eigenvectors of the covariance matrix. It is well-known that the eigenstructure of the sample covariance matrix is a highly distorted version of the population eigenstructure unless the ratio p/n of the number of assets to the number of observations is very small. Indeed, the largest sample eigenvalues are larger than they should be while the smallest ones are smaller. As an example, the range of the distribution of the eigenvalues of the sample covariance matrix derived from n iid random vectors in \mathbb{R}^p whose entries are independent standard Gaussian random variables varies between the two bounds given by $\left(1 \pm \sqrt{p/n}\right)^2$, in the limit of large n and p , instead of a mass at the point 1. Hence, bias corrections are necessary in order to squeeze the sample spectrum and make it closer to the population one. It is the seminal idea introduced by Stein (1975), Haff (1991) or more recently by El Karoui (2008). As a consequence, we will not consider these approaches in the paper and we will just restrict to a standard PCA.

3.2 The shrinkage approach

The shrinkage approach is based on the minimization of a quadratic loss function. It was originally introduced by Stein (1956) and provides an optimal mix between the sample estimate of the covariance/precision matrix and a target matrix. More recently, Ledoit and Wolf (2004a, 2004b, 2004c) considered the shrinkage toward a scalar matrix and toward the covariance matrix implied by Sharpe’s market model while Bengtsson and Holst (2003) considered shrinking the sample covariance matrix toward the covariance matrix derived from a latent factors model estimated by principal component analysis¹. The loss function they consider is the Mean Squared Error and the optimal shrinkage parameter is such that it achieves the best trade-off between the bias and the variance of the resulting estimator. All in all, Disatnik and Benninga (2007) suggest that the simplest approach to shrinkage provides the best results. However the recent advances proposed by Chen *et al.* (2010, 2011) show that better approximations of the covariance matrix can be obtained on the basis of improved shrinkage parameters in particular in the case where in the input data are fat-tailed. Alternatively, non-linear shrinkage methods either based on the introduction of an upper limit for the condition number of the estimated covariance matrix (Won and Kim 2006, Tanaka and Nakata 2013) or on the Marcenko and Pastur (1967) equation seem to provide significant improvements (Ledoit and P ech e 2011, Ledoit and Wolf 2012).

Nonetheless, we restrict our attention to the case of linear shrinkage and use the Oracle Approximating Shrinkage (OAS) estimator introduced by Chen *et al.* (2010) for the shrinkage toward the identity matrix since its performance is actually very close to the performance of the oracle shrinkage estimator, i.e., the estimator obtained when the population parameters are known.

Lemma 1 (Chen *et al.* 2010, Theorem 3). *Under the assumption of iid normally distributed asset returns, given the unbiased sample covariance matrix estimator S_n , the Oracle Approximating Shrinkage estimator of the covariance matrix toward the identity matrix is*

$$\hat{\Sigma}_{OAS} = \hat{\rho}_{OAS} \cdot \frac{\text{Tr } S_n}{p} \cdot \text{Id}_p + (1 - \hat{\rho}_{OAS}) \cdot S_n, \quad (12a)$$

$$\hat{\rho}_{OAS} = \min \left\{ \frac{\left(1 - \frac{2}{p}\right) \text{Tr}(S_n^2) + (\text{Tr } S_n)^2}{\left(n - \frac{2}{p}\right) \cdot \left[\text{Tr}(S_n^2) - \frac{(\text{Tr } S_n)^2}{p}\right]}, 1 \right\}. \quad (12b)$$

The proof of this result is recalled in Appendix C as well as a generalization to the shrinkage toward a diagonal matrix with free diagonal parameters which, to the best of our knowledge, is a new result.

¹The shrinkage parameter for several classical models can be found in Sch afer and Strimmer (2005).

The shrinkage approach can also be successfully applied to the estimation of the precision matrix, which may be more relevant than the application of the shrinkage to the covariance matrix itself since, as recalled in section 2, the solution to the mean-variance optimization program directly involves this former one. When the sample covariance matrix is well-conditioned, namely when the number of observations n is larger than the number of assets p , Haff (1979) provides several random shrinkage estimators that outperform the naive estimator obtained by inversion of the sample covariance matrix. The proposed strategy is, in essence, quite close to the strategy applied by Ledoit and Wolf for the shrinkage of the covariance matrix. Now, when the sample covariance matrix is singular, so that the previous method does not apply, Kubokawa and Srivastava (2008) recently provides a shrinkage method to improve on the classical Moore-Penrose generalized inverse.

In the context of portfolio optimization, the shrinkage of the precision matrix has only been recently considered by Kourtis *et al.* (2012) who propose a non-parametric cross-validation method for the estimation of the shrinkage parameter. We depart from their approach and propose a closed-form OAS estimator for the precision matrix when the sample covariance matrix is well-conditioned and its inverse admits a finite second order moment, i.e. $n > p + 4$.

Lemma 2. *Under the assumption of iid normally distributed asset returns, given the unbiased sample precision matrix estimator $P_n = \frac{n-p-2}{n-1} \cdot S_n^{-1}$ with finite second order moment, the Oracle Approximating Shrinkage estimator of the precision matrix toward the identify matrix is*

$$\hat{\Theta}_{OAS} = \hat{\rho}_{OAS} \cdot \frac{\text{Tr } P_n}{p} \cdot \text{Id} + (1 - \hat{\rho}_{OAS}) \cdot P_n, \quad (13a)$$

$$\hat{\rho}_{OAS} = \min \left\{ \frac{\frac{n-p-\frac{2}{p}(n-p-2)}{n-p-1} \cdot \text{Tr}(P_n^2) + \frac{n-p-2-\frac{2}{p}}{n-p-1} \cdot (\text{Tr } P_n)^2}{\left[\frac{n-p-\frac{2}{p}(n-p-2)}{n-p-1} + n - p - 4 \right] \cdot \left[\text{Tr}(P_n^2) - \frac{(\text{Tr } P_n)^2}{p} \right]}, 1 \right\}. \quad (13b)$$

We postpone the proof to Appendix C which also provides a generalization of this result to the shrinkage toward a diagonal matrix.

The singular case is much more tedious to handle since the sample covariance matrix does not admit a regular inverse. The sample precision matrix is then usually estimated by help of the Moore-Penrose generalized inverse whose statistic follows the generalized inverse Wishart distribution (Bodnar and Okhrin 2008). Unfortunately, to the best of our knowledge, the moments of this distribution do not admit known closed-form expressions apart from the case of cross-sectionally uncorrelated returns (Cook and Forzani 2011). Hence the derivation of the shrinkage estimator of the precision matrix in the singular case is left to future works.

3.3 The sparsity approach

The celebrated principle of parsimony (Occam's razor) has also long ago been summoned for large covariance or precision matrices estimation (Dempster 1972, Dahl *et al.* 2008, Friedman *et al.* 2008, Boyd *et al.* 2010, Cai *et al.* 2011, Lian 2011, e.g.). Such a calling upon parsimony in that context may be motivated from two different origins: Either from an *a priori* sparse modeling choice or from estimation issues.

Assuming *a priori* a sparse dependence model, i.e., the fact that, beyond the diagonal terms, only a small (compared to $p(p - 1)$) number of entries of covariance or precision matrices theoretically differ from zero may first stem from some *theoretical* or *background* knowledge on the system governing the data at hand: Assets belonging to a given class shall be related together while assets pertaining to different classes are more likely to be independent. It then remains an open and difficult question to decide whether such a relative *independence* of classes of assets is better modeled with non diagonal zeroed entries in the covariance or in the precision matrix. When the covariance matrix is chosen sparse, its corresponding inverse, the precision matrix, is usually not sparse (and vice-versa). As a consequence, assuming that either the covariance or the precision matrix is sparse amounts to choosing from the very beginning between two different structural models. Sparse covariance is equivalent, in a Gaussian framework, to consider that the corresponding covariates are independent. It is likely more relevant when one considers assets traded on different markets with weak cross-market correlations, thus yielding block-sparse covariance matrices. Conversely, sparse precision corresponds, within that same framework, to covariates that are conditionally independent. It thus appears more naturally when assets returns can be assumed to be linearly related, so that given the knowledge of a given subset, the remainders are uncorrelated. Beyond, these theoretical considerations, the numerical experimentations and analyses reported in Section 4 below can be read as elements of answers, in the context of practical portfolio allocation performance, to the challenging issue of deciding between a sparse *a priori* imposed to covariance or precision.

The second category of reasons motivating sparsity in dependence matrices stems from the well-known *screening* effect that accompanies large covariance or precision matrix estimation (Hero and Rajaratnam 2011): For large matrices, estimated from short sample size, i.e., when $n \gtrsim p$ or even when $n \lesssim p$, estimation performance for the non diagonal entries are such that it cannot be decided whether small values correspond to actual non zero correlations or to estimation fluctuations, and thus noise. Therefore, small values should be discarded and large values only are significantly estimated and should be further used.

In both cases – sparse modeling or estimation issues – the practical challenge is to decide how many and which non diagonal entries should be set to zero. There have been on-going

efforts to address sparse matrix estimation issues, concentrating first on the precision matrix (Dempster 1972, Dahl *et al.* 2008, Friedman *et al.* 2008, Boyd *et al.* 2010, Cai *et al.* 2011, Lian 2011) and more recently on the covariance matrix (Bien and Tibshirani 2011, Rothman 2012). In essence, the estimation of sparse precision matrices relies on minimizing a cost function, consisting of a balance between a data fidelity term associated to the precision matrix and a penalty term aiming at promoting sparsity. A state-of-the-art formulation of this problem is now referred to as the *Graphical Lasso* (Friedman *et al.* 2008). It balances the negative log-likelihood function, thus relying on the Graphical Gaussian Model framework, and hence following the original formulation due to Dempster (1972), with an l_1 penalization of the estimated precision matrix:

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \operatorname{Tr}(S_n \Theta) - \log \det \Theta + \lambda \cdot \|\Theta\|_1, \quad (14)$$

where λ denotes a penalization parameter to be selected. Indeed, l_1 penalization has been observed to act as an efficient surrogate of l_0 penalization, that explicitly counts non zero entries, yet results in a non convex optimization problem. Instead, estimating Θ from Eq. (14) thus amounts to solving a convex optimization problem, and practical solutions were described in the literature, the two most popular relying on the so-called *path-wise coordinate descent* (Bien and Tibshirani 2011) or *Alternating Direction Method of Multipliers* algorithms (Boyd *et al.* 2010). In the present contribution, used is made of this latter algorithm.

Sparsity can be imposed onto the covariance matrix through the same formulation:

$$\hat{\Sigma} = \underset{\Sigma}{\operatorname{argmin}} \operatorname{Tr}(S_n \Sigma^{-1}) + \log \det \Sigma + \lambda \cdot \|\Sigma\|_1, \quad (15)$$

which however consists of a non-convex problem and is hence far more difficult to solve. It has however been observed that the argument in Eq. (15) can actually be split into concave and a convex function, and that minimization can thus be performed by a majorization-minimization algorithm (Bien and Tibshirani 2011). The corresponding procedure has kindly been made available to us by the authors of (Bien and Tibshirani 2011) and is used in Section 4.

4 Empirical results

We now turn to the implementation of the different portfolio optimization strategies on a dataset made of the daily returns of the $p = 211$ stocks comprised in the Euro Stoxx 600 index between December 14, 2001 and January 24, 2013. We consider both the constrained and the unconstrained optimization programs exposed in Section 2 (see Appendix D for details on the optimization algorithms in the presence of short-sale restrictions) and different estimators of

the covariance and the precision matrices. The covariance and precision matrices are estimated over three rolling windows of size $n = 150, 200$ and 350 days. In the first case, the sample covariance matrix is a highly singular matrix, in the second one a matrix near its critical point and in the last one a full rank matrix. The portfolios, whose inception date is April 15, 2003, are rebalanced every week (five trading days). We do not account for transaction costs. The results are gathered in Tables 1 to 3 which report the out-of-sample performance of the optimal portfolios.

[Introduce Table 1 somewhere here]

[Introduce Table 2 somewhere here]

[Introduce Table 3 somewhere here]

As detailed in the previous section, we consider four classes of estimators of the covariance and precision matrices. The first group of estimators focuses on the usual sample covariance estimator, the identity matrix which yields the equally-weighted portfolio and the diagonal matrix of the sample variances. For the sample covariance matrix, the Moore-Penrose generalized inverse is used when necessary, i.e, for rolling windows 150 and 200 days. We do not report the results for the precision matrix since, within this group, the precision matrix estimator is (up a multiplicative factor that ensures the unbiasedness) the inverse of the corresponding covariance matrix estimator. Hence the optimization results are the same in both cases.

The second group of estimators pertains to the class of (latent) factor models. We consider the estimators derived from a Principal Component Analysis with one and two factors, and an estimator derived from the single index market model, with the Euro Stoxx 600 index as a proxy for the market factor. In this later case, we use unrelated estimators of the covariance and the precision matrix, i.e., the estimator of the precision matrix is not merely the inverse of the covariance matrix estimator (see section 3.1). For the PCA-based estimators, the derivation of specific precision matrix estimators is left for future developments.

The third group of estimators belongs to the class of shrinkage estimators. We consider the shrinkage of the sample covariance and precision matrices toward the identity matrix (times an overall scaling factor) and toward the diagonal matrix of the variances or precisions of each stock. The shrinkage of the precision matrix is based on the method derived in section 3.2. For the time being, it requires the sample covariance matrix to be invertible. Hence, we only report the results obtained with a shrunked precision matrix for the rolling window of 350 days.

Finally, the fourth class of estimators is based upon the sparsity approach exposed in section 3.3 and we report both the results obtained by use of the covariance matrix and the precision matrix. In each case, we only report the results for the value of the penalization parameter λ

which leads to the smallest out-of-sample variance of the GMVP.

In terms of out-of-sample standard deviation, the shrunked covariance matrices outperform the other covariance-based methods both with or without short-sale restrictions. Nonetheless the superiority of the shrinkage approach diminishes with the size of the rolling window and is lower in the presence of constraints. As for the precision-based approaches, both the sparsity and PCA provide the best results, depending on the length of the rolling window under consideration. Again, the differences between the methods tend to fade away with the length of the rolling window and the presence of constraints. But overall, at the noticeable exception of the sparse precision matrix for the rolling window of 150 days, the shrinkage of the covariance matrix always leads to the GMVP with the lowest out-of-sample standard deviation among all the estimation methods we have implemented in this study.

Now, regarding the performance in terms of Sharpe ratio, the shrinkage of the covariance matrix toward the identity matrix uniformly dominates of the covariance-based strategies irrespective of the size of the rolling window and of the presence of constraints or not. We also notice that for the shortest rolling windows (150 and 200 days) for which for the number of assets is larger than the sample size, the constrained optimization program leads to GMVP with higher Sharpe ratios compared to the unconstrained case. This observation holds not only for the best performing strategy, but for all the covariance-based strategies (at the exception of the PCA approach with two latent factors). The reverse phenomenon is observed for the longest rolling window (350 days). Hence, in line with Jagannathan and Ma (2003), we can conclude that the introduction of constraints makes the optimization process more robust when the sample covariance matrix is ill-conditioned.

For the precision-based approaches, the sparse matrix estimation method always leads to the highest Sharpe ratios in the absence of short-sale restrictions and still dominates in the presence of restrictions for the longest rolling window. We do not observe that the introduction of constraints help stabilize the optimization process for small sample size. On the contrary, on average, the introduction of constraints spoils the performance of the GMVP. Besides when we compare the best precision-based strategy to the best covariance-based strategy, we can only conclude to the lack of a significant difference. Thus, depending on the chosen approach, either based on the estimation of the covariance or based on the precision matrix, the champion of each group is different but they lead to comparable *ex-post* performance.

However, if we look at the average composition of the optimal portfolios, divergences appear between the best performing covariance and precision-based methods. For the 150 days rolling window, if we compare the best performing GMVP in the absence of constraints based on both methods, we observe that the concentrations of the portfolios, as measured by the Herfindal

index (Woerheide and Persson 1993), remain close one to the other, but the turnover and the short interest are more than three times smaller for the precision-based portfolio, which can be considered as a significant advantage in a practical perspective. The picture remains the same for all rolling windows in the absence of short-sale restrictions. The composition of the best performing portfolio derived from the precision matrix is more stable (in terms of turnover) and exhibits lower levels of short interests. In addition, while the best performing GMVP based on the covariance approach remains concentrated on a dozen of stocks irrespective of the size of the rolling window, the best performing GMVP based on the precision approach becomes less concentrated as the size of the rolling window increases. Of course, the introduction of short-sale restrictions naturally enforces a high-level of diversification, hence lowers the concentration and the turnover of the portfolio, and thus rubs out the differences between the two approaches according to these criteria.

To sum up, our empirical results show that in terms of risk only, as measured by the out-of-sample standard deviation of the returns of the GMVP, the shrinkage of the covariance matrix dominates irrespective of the sample size and of the presence or the absence of short-sale restrictions. Nonetheless, this lower level of out-of-sample risk does not translate into a better out-of-sample performance, as measured by the *ex-post* Sharpe ratio of the GMVP. Actually, we do not observe any significant difference between the performance of the champions of the covariance and the precision-based approaches. However, in the absence of constraints, the precision-based approach leads to more stable GMVPs, as measured by lower concentration, lower turnover rate and lower level of short interests. Hence, as far as performance as well as risk are concerned, the direct estimation of the precision matrix, and in particular of sparse precision matrices, seems relevant for the construction of optimal portfolios from a large universe of stocks.

5 Conclusion

We have conducted an in-depth study of the relative performance of different estimation strategies of the GMVP based of the inversion of the estimated covariance matrix or the direct estimation of the precision matrix. Our first contribution is the confirmation of the relevance of the shrinkage and the positive impact of the constraints on the *ex-post* performance of the GMVP when focusing on the covariance-based optimization approach. However, if we confirm the superiority of the equally-weighted portfolio in terms of raw return, it is not the case on a risk adjusted basis. More important, our analysis of the precision-based approach shows that, if the gain in terms of Sharpe ratio is not significant with respect to the standard covariance-based approach, the former leads to much more stable optimal portfolios even in the absence of

additional, and sometimes irrelevant, constraints.

We think that these empirical results are of interest both from an academic and a professional point of view in so far as they pave the way toward the development of new estimation methods for optimal portfolio weights in the mean-variance framework and yield new questions regarding the informational content of the sample covariance and precision matrices.

References

- Bai, J., and S. Ng (2002) Determining the Number of Factors in Approximate Factor Models. *Econometrica*, **70**, 191–221.
- Bai, J., and S. Shi (2011) Estimating High Dimensional Covariance Matrices and its Applications. *Annals of Economics and Finance*, **12**, 199–215.
- Bajeux-Besnainou, I., W. Bandara, and E. Bura (2012) A Krylov Subspace Approach to Large Portfolio Optimization. *Journal of Economic Dynamics & Control* **36**, 1688–1699.
- Bengtsson, C., and J. Holst (2003) On Portfolio Selection: Improved Covariance Matrix Estimation for Swedish Asset Returns. *Working paper* Lund University.
- Bien, J., and R. J. Tibshirani (2011) Sparse Estimation of a Covariance Matrix. *Biometrika* **98**, 807–820.
- Bodnar, T., and Y. Okhrin (2008) Properties of the Singular, Inverse and Generalized Inverse Partitioned Wishart Distributions. *Journal of Multivariate Analysis* **99**, 2389–2405.
- Boyd, S., N. Parikh, E. Chu, B. Peleato and J. Eckstein (2010) Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends in Machine Learning* **3**, 1–122.
- Cai, T., W. Liu, and X. Luo (2011) A Constrained l_1 Minimization Approach to Sparse Precision Matrix Estimation. *Journal of the American Statistical Association* **106**, 594–607.
- Carhart, M. M. (1997) On Persistence in Mutual Fund Performance. *Journal of Finance* **52**, 57–82.
- Chamberlain, G., and M. Rothschild (1983) Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets. *Econometrica* **51**, 1305–1324.
- Chen, Y., A. Wiesel, Y. C. Eldar and A. O. Hero (2010) Shrinkage Algorithms for MMSE Covariance Estimation. *IEEE Transactions on Signal Processing* **58**, 5016–5029.

- Chen, Y., A. Wiesel and A. O. Hero (2011) Robust shrinkage estimation of high dimensional covariance matrices. *IEEE Transactions on Signal Processing* **59**, 4097–4107.
- Connor, G., (1982) Asset pricing in factor economies. *Doctoral dissertation* (Yale university).
- Cook, R. D., and L. Forzani (2011) On the Mean and Variance of the Generalized Inverse of a Singular Wishart Matrix. *Electronic Journal of Statistics* **5**, 146–158.
- Dahl, J., L. Vandenberghe and V. Roychowdhury (2008) Covariance selection for non-chordal graphs via chordal embedding. *Optimization Methods and Software* **23**, 501–520.
- DeMiguel, V., L. Garlappi and R. Uppal (2009) Optimal versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *Review of Financial Studies* **22**, 1915–1953.
- Dempster, A. P. (1972) Covariance Selection. *Biometrics* **28**, 157–75
- Desmoulin-Lebeault, F., and C. Kharoubi (2012) Non-Gaussian Diversification: When Size Matters. *Journal of Banking and Finance* **36**, 1987-1996.
- Disatnik, D. J., and S. Benninga (2007) Shrinking the Covariance Matrix – Simpler is Better. *Journal of Portfolio Management* **33(4)**, 55–63.
- El Karoui, N. (2008) Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Annals of Statistics*, **36**, 2757–2790.
- Elton, E. J., and M. J. Gruber (1973) Estimating the Dependence Structure of Share Prices – Implications for Portfolio Selection, *Journal of Finance* **28**, 1203–1232.
- Evans, J., and S. Archer (1968) Diversification and the Reduction of Dispersion: An Empirical analysis. *Journal of Finance* **23**, 761–767.
- Fama, E. F., and K. R. French (1992) The Cross-Section of Expected Stock Returns. *Journal of Finance* **47**, 427–465.
- Friedman, J., H. Trevor, and R. Tibshirani (2008) Sparse Inverse Covariance Estimation with the Graphical Lasso. *Biostatistics* **9**, 432–441.
- Haff, L. R. (1979) Estimation of the Inverse Covariance Matrix: Random Mixtures of the Inverse Wishart Matrix and the Identity. *Annals of Statistics* **7**, 1264-1276.
- Haff, L. R. (1991) The Variational Form of Certain Bayes Estimators. *Annals of Statistics* **19**, 1163–1190.

- Hero, A. O., and B. Rajaratnam (2011) Large Scale Correlation Screening. *Journal of the American Statistical Association* **106**, 1540–1552.
- Jagannathan, R., and T. Ma (2003) Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. *Journal of Finance* **58**, 1651-1684.
- Kourtis, A., G. Dotsis, and R. N. Markellos (2012) Parameter Uncertainty in Portfolio Selection: Shrinking the Inverse Covariance Matrix. *Journal of Banking and Finance* **36**, 2522–2531.
- Kubokawa, T. and M. S. Srivastava (2008) Estimation of the Precision Matrix of a Singular Wishart Distribution and its Application in High-Dimensional Data. *Journal of Multivariate Analysis* **99**, 1906–1928.
- Laloux, L., P. Cizeau, J. P. Bouchaud, and M. Potters (1999). Noise Dressing of Financial Correlation Matrices. *Physical Review Letters* **83**, 1467-1470.
- Laloux, L., P. Cizeau, J. P. Bouchaud, and M. Potters 2000. Random Matrix Theory and Financial Correlations. *International Journal of Theoretical and Applied Finance* **3**, 391–397.
- Ledoit, O. and S. Péché (2011) Eigenvectors of some large sample covariance matrix ensembles. *Probability Theory and Related Fields* **151**, 233–264.
- Ledoit, O., and M. Wolf (2004a) A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices. *Journal of Multivariate Analysis* **88**, 365-411.
- Ledoit, O., and M. Wolf (2004b) Improved Estimation of the Covariance Matrix of Stock Returns With an Application to Portfolio Selection. *Journal of Empirical Finance* **10**, 603–621.
- Ledoit, O., and M. Wolf (2004c) Honey, I Shrunk the Sample Covariance Matrix. *Journal of Portfolio Management* **30(4)**, 110–119.
- Ledoit, O., and M. Wolf (2012) Nonlinear Shrinkage Estimation of Large-Dimensional Covariance Matrices. *Annals of Statistics* **40**, 1024–1060.
- Lian, H. (2011) Shrinkage Tuning Parameter Selection in Precision Matrices Estimation. *Journal of Statistical Planning and Inference* **141**, 2839–2848.
- Malevergne, Y., and D. Sornette (2004) Collective Origin of the Coexistence of Apparent Random Matrix Theory Noise and of Factors in Large Sample Correlation matrices. *Physica A* **331**, 660–668.

- Marcenko, V. A., and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Sbornik: Mathematics* **1**, 457–483
- Michaud, R. O. (1989) The Markowitz Optimization Enigma: Is ‘Optimized’ Optimal? *Financial Analysts Journal* **45**, 31–42.
- Muirhead, R. J. (2005) *Aspects of Multivariate Statistical Theory* (Wiley-Interscience, 2nd edition)
- Pantaleo, E., M. Tumminello, F. Lillo and R. N. Mantegna (2011) When do Improved Covariance Matrix Estimators Enhance Portfolio Optimization? An Empirical Comparative Study of Nine Estimators. *Quantitative Finance* **11**, 1067–1080.
- Rothman, A. J. (2012) Positive definite estimators of large covariance matrices. *Biometrika* **99**, 733–740.
- Schäfer, J. and K. Strimmer (2005) A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics. *Statistical Applications in Genetics and Molecular Biology* **4**, DOI: 10.2202/1544-6115.1175.
- Sharpe, W. (1963) A Simplified Model for Portfolio Analysis. *Management Science* **79**, 277–231.
- Siskind, V. (1972) Second Moments of Inverse Wishart-Matrix Elements. *Biometrika* **59**, 690–691.
- Statman, M. (1987) How Many Stocks Make a Diversified Portfolio. *Journal of Financial and Quantitative Analysis* **22**, 353–363.
- Stein, C. (1956) Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* **1**, 197–206.
- Stein, C. (1975) Estimation of a Covariance Matrix. *Rietz Lecture, 39th Annual Meeting IMS*. Atlanta, Georgia.
- Tanaka, M., and K. Nakata (2013) Positive definite matrix approximation with condition number constraint. *Optimization Letters*, DOI:10.1007/s11590-013-0632-7.
- Wigner, E. P. (1953) On a Class of Analytic Functions from the Quantum Theory of Collisions. *Annals of Mathematics* **53**, 36–67.
- Woerheide, W., and D. Persson (1993) An Index of Portfolio Diversification. *Financial Review Services* **2**, 73–85.

Won, J. H., and S. J. Kim (2006) Maximum likelihood covariance estimation with a condition number constraint. *Fortieth Asilomar Conference on Signals, Systems, and Computers*, 1445–1449

Yang, R. and J. O. Berger (1994) Estimation of a Covariance Matrix Using the Reference Prior. *The Annals of Statistics* **22**, 1195–1211.

A Estimation of the precision matrix in the constant correlation model

Under the assumption of normally distributed assets returns with values $\mu \in \mathbb{R}^p$ and covariance matrix Σ , the log-likelihood of the precision matrix $\Theta = \Sigma^{-1}$ reads

$$\log \det \Theta - \text{Tr}(\Theta S_n), \quad (16)$$

where S_n denotes the sample covariance matrix estimate from n iid random vectors of assets returns.

Under the constant correlation assumption, the covariance matrix reads $\Sigma = \Delta R \Delta$ where $\Delta = \text{diag}(\sigma_1, \dots, \sigma_p)$ and R is the correlation matrix

$$R = (1 - \rho)Id + \rho \mathbf{1}\mathbf{1}^t, \quad (17)$$

with $\rho > -1/(p-1)$. By application of the Sherman-Morrison inversion formula, we get

$$R^{-1} = \frac{1}{1 - \rho} \left[Id - \frac{\rho}{1 + (p-1)\rho} \mathbf{1}\mathbf{1}^t \right]. \quad (18)$$

Thus, given that the eigenvalues of R are $1 + (p-1)\rho$ and $1 - \rho$, with multiplicity $p-1$, we get

$$\log \det \Theta = \log \det \Delta^{-1} R^{-1} \Delta^{-1}, \quad (19)$$

$$= 2 \sum_{i=1}^p \log \sigma_i^{-1} - \log [1 + (p-1)\rho] - (p-1) \log(1 - \rho) \quad (20)$$

while

$$\text{Tr}(\Theta S_n) = \text{Tr}[\Delta^{-1} R^{-1} \Delta^{-1} S_n], \quad (21)$$

$$= \text{Tr}[R^{-1} \Delta^{-1} S_n \Delta^{-1}], \quad (22)$$

$$= \frac{1}{1 - \rho} \text{Tr}[\Delta^{-1} S_n \Delta^{-1}] - \frac{\rho}{(1 - \rho)[1 + (p-1)\rho]} \text{Tr}[\mathbf{1}\mathbf{1}^t \Delta^{-1} S_n \Delta^{-1}], \quad (23)$$

$$= \frac{1}{1 - \rho} \text{Tr}[\Delta^{-1} S_n \Delta^{-1}] - \frac{\rho}{(1 - \rho)[1 + (p-1)\rho]} \mathbf{1}^t \Delta^{-1} S_n \Delta^{-1} \mathbf{1}. \quad (24)$$

Finally, the log likelihood reads

$$\begin{aligned} \log L(\{\sigma_i\}_{i=1}^p, \rho) &= \sum_{i=1}^p \log \sigma_i^{-1} - \log [1 + (p-1)\rho] - (p-1) \log(1-\rho) \\ &\quad - \frac{1}{1-\rho} \text{Tr} [\Delta^{-1} S_n \Delta^{-1}] \\ &\quad + \frac{\rho}{(1-\rho)[1+(p-1)\rho]} \mathbf{1}^t \Delta^{-1} S_n \Delta^{-1} \mathbf{1}. \end{aligned} \quad (25)$$

B The one factor model

We consider the one factor model

$$r_t = \alpha + \beta \cdot r_{m,t} + \varepsilon_t, \quad (26)$$

where r_t denotes the p -vector of assets returns on day $t \in \{1, \dots, n\}$, $r_{m,t}$ is the return on the market index on day t , α and β are the p -vectors of intercepts and factor loadings while ε_t is the p -vector of residuals with diagonal covariance matrix Δ .

The covariance and precision matrices then reads

$$\Sigma = \Delta + \sigma_m^2 \beta \beta', \quad (27)$$

The OLS estimate of the factor loading for asset $i \in \{1, \dots, p\}$ satisfies

$$\hat{\beta}_i = \beta_i + \frac{\mathbf{1}'_n (1_n \cdot r'_m - r_m \cdot \mathbf{1}'_n) \varepsilon_i}{n \cdot r'_m r_m - (\mathbf{1}'_n r_m)^2}, \quad (28)$$

where $\mathbf{1}_n$ is the n -vector of ones, from which we immediately get

$$\text{E} \left[\hat{\beta}_i \hat{\beta}_j | r_m \right] = \beta_i \beta_j + \frac{n \cdot \Delta_{ij}}{n \cdot r'_m r_m - (\mathbf{1}'_n r_m)^2}. \quad (29)$$

Hence, by the law of iterated expectations, given the unbiased estimator of σ_m^2

$$\hat{\sigma}_m^2 = \frac{1}{n-1} \sum_{t=1}^n \left(r_{m,t} - \frac{1}{n} \sum_{s=1}^n r_{m,t} \right)^2, \quad (30)$$

$$= \frac{n \cdot r'_m r_m - (\mathbf{1}'_n r_m)^2}{n \cdot (n-1)}, \quad (31)$$

we obtain, under normality,

$$\mathbb{E} \left[\hat{\sigma}_m^2 \cdot \hat{\beta} \hat{\beta}' \right] = \mathbb{E} \left[\hat{\sigma}_m^2 \cdot \mathbb{E} \left[\hat{\beta} \hat{\beta}' | r_m \right] \right] = \sigma_m^2 \cdot \beta \beta' + \frac{1}{n-1} \Delta. \quad (32)$$

Thus, considering the unbiased estimator of Δ , i.e. the sum of the squared empirical residuals normalized by $(n-2)$, we conclude that

$$\hat{\Sigma} = \frac{n-2}{n-1} \cdot \hat{\Delta} + \hat{\sigma}_m^2 \hat{\beta} \hat{\beta}', \quad (33)$$

is an unbiased estimator of the covariance matrix.

As for the precision matrix, which reads

$$\Theta = \Delta^{-1} - \frac{\Delta^{-1} \beta \cdot \beta' \Delta^{-1}}{\frac{1}{\sigma_m^2} + \beta' \Delta^{-1} \beta}, \quad (34)$$

or equivalently

$$\Theta = \Delta^{-1} - \frac{\sigma_m^2 \cdot \Delta^{-1} \beta \cdot \beta' \Delta^{-1}}{1 + \sigma_m^2 \cdot \beta' \Delta^{-1} \beta}, \quad (35)$$

we can improve on the basic plug-in estimator if we account for the bias of $\hat{\Delta}^{-1}$

$$\mathbb{E} \left[\hat{\Delta}^{-1} \right] = \frac{n-2}{n-4} \cdot \Delta^{-1} \quad (36)$$

since each diagonal entry of $\hat{\Delta}$ follows a χ^2 -distribution with $n-2$ degrees of freedom.

In addition, under normality, the estimators of β and Δ are independent, hence

$$\mathbb{E} \left[\frac{\hat{\beta}_i \cdot \hat{\beta}_j}{\Delta_{ii} \cdot \Delta_{jj}} \middle| r_m \right] = \mathbb{E} \left[\hat{\beta}_i \cdot \hat{\beta}_j | r_m \right] \cdot \mathbb{E} \left[\Delta_{ii}^{-1} | r_m \right] \cdot \mathbb{E} \left[\Delta_{jj}^{-1} | r_m \right], \quad (37)$$

$$= \left(\beta_i \beta_j + \frac{n \cdot \Delta_{ij}}{n \cdot r_m' r_m - (1_n' r_m)^2} \right) \cdot \left(\frac{n-2}{n-4} \right)^2 \cdot \frac{1}{\Delta_{ii} \cdot \Delta_{jj}}, \quad (38)$$

$$= \left(\frac{n-2}{n-4} \right)^2 \cdot \left[\frac{\beta_i \cdot \beta_j}{\Delta_{ii} \cdot \Delta_{jj}} + \frac{1}{(n-1) \cdot \hat{\sigma}_m^2} \cdot \frac{\Delta_{ij}}{\Delta_{ii} \cdot \Delta_{jj}} \right], \quad (39)$$

so that, by the law of iterated expectations, we obtain

$$\mathbb{E} \left[\hat{\sigma}_m^2 \cdot \hat{\Delta}^{-1} \hat{\beta} \cdot \hat{\beta}' \hat{\Delta}^{-1} \right] = \left(\frac{n-2}{n-4} \right)^2 \cdot \left[\sigma_m^2 \cdot \Delta^{-1} \beta \cdot \beta' \Delta^{-1} + \frac{1}{(n-1)} \cdot \Delta^{-1} \right] \quad (40)$$

Similarly, we can show that

$$\mathbb{E} \left[1 + \hat{\sigma}_m^2 \cdot \hat{\beta}' \hat{\Delta}^{-1} \hat{\beta} \right] = 1 + \frac{n-2}{n-4} \left[\sigma_m^2 \cdot \beta' \Delta^{-1} \beta + \frac{p}{n-1} \right], \quad (41)$$

hence an improved estimator of the precision matrix is

$$\hat{\Theta} = \frac{n-4}{n-2} \hat{\Delta}^{-1} + \frac{\left(\frac{n-4}{n-2}\right)^2 \hat{\sigma}_m^2 \hat{\Delta}^{-1} \hat{\beta} \hat{\beta}' \hat{\Delta}^{-1} - \frac{1}{n-1} \hat{\Delta}^{-1}}{1 + \frac{n-4}{n-2} \hat{\sigma}_m^2 \hat{\beta}' \hat{\Delta}^{-1} \hat{\beta} - \frac{p}{n-1}}. \quad (42)$$

C The shrinkage approach

We follow the approach of Ledoit and Wolf (2004b) and, in a first time, we apply it to the derivation of the optimal shrinkage toward the identity matrix. We are thus look for the parameter ρ and ν which minimizes the quadratic loss function

$$L(\nu, \rho) = \mathbb{E} \left[\|\rho \nu Id + (1 - \rho) S_n - \Sigma\|^2 \right] \quad (43)$$

where $\|\cdot\|$ denotes the Frobenius norm while S_n denotes the unbiased sample covariance matrix estimate from n iid random vectors of Gaussian assets returns with covariance matrix Σ .

The expansion of the quadratic loss function reads

$$L(\nu, \rho) = \rho^2 \text{Tr} (\nu Id - \Sigma)^2 + (1 - \rho)^2 \mathbb{E} \left[\text{Tr} (S_n - \Sigma)^2 \right], \quad (44)$$

since S_n is an unbiased estimator of Σ so that the cross-product cancels out. The differentiation with respect to ν yields

$$\nu = \frac{1}{p} \text{Tr} \Sigma, \quad (45)$$

i.e., ν is the average variance, that will be estimated as follows

$$\hat{\nu} = \frac{1}{p} \text{Tr} S_n. \quad (46)$$

By substitution into the quadratic loss function we get

$$\begin{aligned} L(\hat{\nu}, \rho) &= \rho^2 \mathbb{E} \left[\text{Tr} \left(\frac{1}{p} (\text{Tr} S_n) Id - \Sigma \right)^2 \right] + 2\rho(1 - \rho) \mathbb{E} \left[\frac{1}{p} \text{Tr} (S_n) \cdot \text{Tr} (S_n - \Sigma) \right] \\ &+ (1 - \rho)^2 \mathbb{E} \left[\text{Tr} (S_n - \Sigma)^2 \right], \end{aligned} \quad (47)$$

and the differentiation with respect to ρ leads, after simple algebraic manipulations, to

$$\rho = \frac{\mathbb{E}[\text{Tr}(S_n^2)] - \text{Tr}(\Sigma^2) - \frac{1}{p}\text{Var}(\text{Tr}S_n)}{\mathbb{E}[\text{Tr}(S_n^2)] - \frac{1}{p}(\text{Tr}\Sigma)^2 - \frac{1}{p}\text{Var}(\text{Tr}S_n)}. \quad (48)$$

Notice that, up to now, this derivation is totally free from the distributional properties of the sample matrix S_n apart from the absence of bias.

Let us now use the fact that the sample covariance matrix follows a Wishart distribution

$$(n-1) \cdot S_n \sim \mathcal{W}_p(n-1, \Sigma), \quad (49)$$

so that (Muirhead 2005)

$$\text{Cov}\left((S_n)_{ij}, (S_n)_{kl}\right) = \frac{1}{(n-1)^2} \cdot \text{Cov}\left((n-1) \cdot (S_n)_{ij}, (n-1) \cdot (S_n)_{kl}\right), \quad (50)$$

$$= \frac{1}{n-1} (\Sigma_{ik} \cdot \Sigma_{jl} + \Sigma_{il} \cdot \Sigma_{jk}). \quad (51)$$

As a consequence

$$\mathbb{E}[\text{Tr}(S_n^2)] = \sum_{i=1}^p \mathbb{E}[(S_n^2)_{ii}] = \sum_{i,j=1}^p \mathbb{E}[(S_n)_{ij}^2], \quad (52)$$

$$= \sum_{i,j=1}^p \left\{ \underbrace{\text{Var}(S_n)_{ij}}_{\frac{\Sigma_{ij}^2 + \Sigma_{ii}\Sigma_{jj}}{n-1}} + \underbrace{\mathbb{E}[(S_n)_{ij}]^2}_{\Sigma_{ij}^2} \right\}, \quad (53)$$

$$= \frac{n}{n-1} \text{Tr}(\Sigma^2) + \frac{1}{n-1} (\text{Tr}\Sigma)^2, \quad (54)$$

and

$$\text{Var}(\text{Tr}S_n) = \text{Var}\left[\sum_{i=1}^p (S_n)_{ii}\right] = \sum_{i,j=1}^p \underbrace{\text{Cov}[(S_n)_{ii}, (S_n)_{jj}]}_{\frac{2}{n-1}\Sigma_{ij}^2}, \quad (55)$$

$$= \frac{2}{n-1} \text{Tr}(\Sigma^2). \quad (56)$$

Notice that these relations can straightforwardly be obtained by application of the Stein-Haff identity.

By substitution of equations (54) and (56) in (48), we obtain

$$\rho = \frac{\left(1 - \frac{2}{p}\right) \text{Tr}(\Sigma^2) + (\text{Tr} \Sigma)^2}{\left(n - \frac{2}{p}\right) \text{Tr}(\Sigma^2) + \left(1 - \frac{n-1}{p}\right) (\text{Tr} \Sigma)^2}. \quad (57)$$

This expression is the same as the one given by equation (7) in Chen *et al.* (2010) with the replacement $n \rightarrow n - 1$ to account for the fact the mean value is unknown in the present case. It can be estimated by the Oracle Approximating Shrinkage (OAS) estimator

$$\hat{\Sigma}_{OAS} = \hat{\rho}_{OAS} \cdot \hat{\nu} \cdot Id + (1 - \hat{\rho}_{OAS}) \cdot S_n, \quad (58)$$

$$\hat{\rho}_{OAS} = \min \left\{ \frac{\left(1 - \frac{2}{p}\right) \text{Tr}(S_n^2) + (\text{Tr} S_n)^2}{\left(n - \frac{2}{p}\right) \cdot \left[\text{Tr}(S_n^2) - \frac{(\text{Tr} S_n)^2}{p}\right]}, 1 \right\}. \quad (59)$$

We can notice, in passing, that there is a typo in eq. (20) of the working paper version of Chen *et al.* (2010); the right equation is

$$\hat{\phi} = \frac{\text{Tr}(S_n^2) - \frac{(\text{Tr} S_n)^2}{p}}{\left(1 - \frac{2}{p}\right) \text{Tr}(S_n^2) + (\text{Tr} S_n)^2}. \quad (60)$$

As for the shrinkage estimator of the precision matrix, the quadratic loss function reads

$$L(\nu, \rho) = E \left[\|\rho \nu Id + (1 - \rho) P_n - \Theta\|^2 \right], \quad (61)$$

where $\Theta = \Sigma^{-1}$ and P_n is the unbiased sample precision matrix obtained by inversion of the sample covariance matrix S_n if $n > p$ or is given by the Moore-Penrose generalized inverse if $n \leq p$.

The same derivations as previously yield

$$\nu = \frac{1}{p} \text{Tr} \Theta, \quad (62)$$

i.e., ν is the inverse of the harmonic mean of the individual variances, which can be estimated by

$$\hat{\nu} = \frac{1}{p} \text{Tr} P_n. \quad (63)$$

As a consequence,

$$L(\hat{\nu}, \rho) = \rho^2 \mathbb{E} \left[\text{Tr} \left(\frac{1}{p} (\text{Tr } P_n) \text{Id} - \Theta \right)^2 \right] + \rho(1 - \rho) \mathbb{E} \left[\frac{1}{p} \text{Tr} (P_n) \cdot \text{Tr} (P_n - \Theta) \right] \\ + (1 - \rho)^2 \mathbb{E} [\text{Tr} (P_n - \Theta)^2], \quad (64)$$

and the differentiation with respect to ρ leads to

$$\rho = \frac{\mathbb{E} [\text{Tr} (P_n^2)] - \text{Tr} (\Theta^2) - \frac{1}{p} \text{Var} (\text{Tr } P_n)}{\mathbb{E} [\text{Tr} (P_n^2)] - \frac{1}{p} (\text{Tr } \Theta)^2 - \frac{1}{p} \text{Var} (\text{Tr } P_n)}. \quad (65)$$

In the case $n > p$, the inverse of the sample covariance exists and the unbiased sample estimator of the precision matrix is

$$P_n = \frac{n - p - 2}{n - 1} \cdot S_n^{-1}. \quad (66)$$

Indeed, given

$$\frac{1}{n - p - 2} \cdot P_n = ((n - 1) \cdot S_n)^{-1} \sim \mathcal{W}_p^{-1}(n - 1, \Theta), \quad (67)$$

where \mathcal{W}^{-1} is the inverse Wishart distribution, we immediately get

$$\mathbb{E} \left[\frac{1}{n - p - 2} \cdot P_n \right] = \frac{1}{n - p - 2} \cdot \Theta, \quad (68)$$

so that P_n is actually unbiased

$$\mathbb{E} [P_n] = \Theta. \quad (69)$$

From Siskind (1972), we know that

$$\text{Cov} \left((P_n)_{ij}, (P_n)_{kl} \right) = (n - p - 2)^2 \cdot \text{Cov} \left(\frac{1}{n - p - 2} \cdot (P_n)_{ij}, \frac{1}{n - p - 2} \cdot (P_n)_{kl} \right) \quad (70)$$

$$= \frac{2 \cdot \Theta_{ij} \cdot \Theta_{kl} + (n - p - 2) (\Theta_{ik} \cdot \Theta_{jl} + \Theta_{il} \cdot \Theta_{jk})}{(n - p - 1)(n - p - 4)}. \quad (71)$$

Hence

$$\mathbb{E} [\text{Tr} (P_n^2)] = \sum_{i=1}^p \mathbb{E} [(P_n^2)_{ii}] = \sum_{i,j=1}^p \mathbb{E} [(P_n)_{ij}^2], \quad (72)$$

$$= \sum_{i,j=1}^p \left\{ \text{Var} (P_n)_{ij} + \mathbb{E} [(P_n)_{ij}]^2 \right\}, \quad (73)$$

$$= \left(\frac{n-p}{(n-p-1)(n-p-4)} + 1 \right) \sum_{i,j=1}^p \Theta_{ij}^2 + \frac{n-p-2}{(n-p-1)(n-p-4)} \sum_{i,j=1}^p \Theta_{ii} \cdot \Theta_{jj}, \quad (74)$$

$$= \left(\frac{n-p}{(n-p-1)(n-p-4)} + 1 \right) \text{Tr} (\Theta^2) \quad (75)$$

$$+ \frac{n-p-2}{(n-p-1)(n-p-4)} (\text{Tr} \Theta)^2, \quad (76)$$

$$(77)$$

and

$$\text{Var} (\text{Tr} P_n) = \text{Var} \left[\sum_{i=1}^p (P_n)_{ii} \right] = \sum_{i,j=1}^p \text{Cov} [(P_n)_{ii}, (P_n)_{jj}], \quad (78)$$

$$= 2 \cdot \sum_{i,j=1}^p \frac{\Theta_{ii} \cdot \Theta_{jj} + (n-p-1) \cdot \Theta_{ij}^2}{(n-p-1)(n-p-4)}, \quad (79)$$

$$= \frac{2 \cdot (\text{Tr} \Theta)^2 + 2(n-p-2) \cdot \text{Tr} (\Theta^2)}{(n-p-1)(n-p-4)}. \quad (80)$$

Thus, the oracle shrinkage parameter is

$$\rho = \frac{\frac{n-p-\frac{2}{p}(n-p-2)}{(n-p-1)(n-p-4)} \cdot \text{Tr} (\Theta^2) + \frac{n-p-2-\frac{2}{p}}{(n-p-1)(n-p-4)} \cdot (\text{Tr} \Theta)^2}{\left[1 + \frac{n-p-\frac{2}{p}(n-p-2)}{(n-p-1)(n-p-4)} \right] \text{Tr} (\Theta^2) + \left[\frac{n-p-2-\frac{2}{p}}{(n-p-1)(n-p-4)} - \frac{1}{p} \right] (\text{Tr} \Theta)^2}. \quad (81)$$

In order to derive an (hopefully) efficient estimator of the oracle shrinkage parameter, we follow the line of Chen *et al.* (2010) and introduce the Oracle Approximating Shrinkage (OAS) estimator for the precision matrix:

Lemma 3 (Shrinkage toward Identity). *The Oracle Approximating Shrinkage estimator of the*

precision matrix Θ toward the identity matrix when $n > p + 2$ is

$$\hat{\Theta}_{OAS} = \hat{\rho}_{OAS} \cdot \frac{1}{p} \text{Tr} P_n \cdot Id + (1 - \hat{\rho}_{OAS}) \cdot P_n, \quad (82)$$

$$\hat{\rho}_{OAS} = \min \left\{ \frac{\frac{n-p-\frac{2}{p}(n-p-2)}{n-p-1} \cdot \text{Tr} (P_n^2) + \frac{n-p-2-\frac{2}{p}}{n-p-1} \cdot (\text{Tr} P_n)^2}{\left[\frac{n-p-\frac{2}{p}(n-p-2)}{n-p-1} + n-p-4 \right] \cdot \left[\text{Tr} (P_n^2) - \frac{(\text{Tr} P_n)^2}{p} \right]}, 1 \right\}. \quad (83)$$

Proof. We follow Chen *et al.* (2010), and define the AOS estimator as the limit of the iterative process

$$\hat{\Theta}_j = \hat{\rho}_j \cdot \frac{1}{p} \text{Tr} P_n \cdot Id + (1 - \hat{\rho}_j) \cdot P_n, \quad (84)$$

and

$$\hat{\rho}_{j+1} = \frac{\left[n-p-\frac{2}{p}(n-p-2) \right] \text{Tr}(\Theta_j P_n) + \left(n-p-2-\frac{2}{p} \right) (\text{Tr} P_n)^2}{\left[n-p-\frac{2}{p}(n-p-2) + (n-p-1)(n-p-4) \right] \text{Tr}(\Theta_j P_n) + \left[n-p-2-\frac{2}{p}-\frac{1}{p}(n-p-1)(n-p-4) \right] (\text{Tr} P_n)^2}. \quad (85)$$

Thus, by substitution of (84) into (85), we get

$$\hat{\rho}_{j+1} = \frac{1 - \left[n-p-\frac{2}{p}(n-p-2) \right] \hat{\phi} \cdot \rho_j}{1 + (n-p-1)(n-p-4) \hat{\phi} - \left[n-p-\frac{2}{p}(n-p-2) - (n-p-1)(n-p-4) \right] \hat{\phi} \cdot \rho_j} \quad (86)$$

with

$$\hat{\phi} = \frac{\text{Tr} (P_n^2) - \frac{(\text{Tr} P_n)^2}{p}}{\left[n-p-\frac{2}{p}(n-p-2) \right] \text{Tr} (P_n^2) + \left(n-p-2-\frac{2}{p} \right) (\text{Tr} P_n)^2}. \quad (87)$$

Hence

$$\lim_{j \rightarrow \infty} \rho_j = \min \left\{ \frac{1}{\left[n-p-\frac{2}{p}(n-p-2) - (n-p-1)(n-p-4) \right] \hat{\phi}}, 1 \right\} \quad (88)$$

by the fixed-point theorem and the result follows immediately. \square

A similar result can be obtained for the shrinkage of the sample precision matrix toward a diagonal matrix:

Lemma 4 (Shrinkage toward diagonal). *The Oracle Approximating Shrinkage estimator of the*

precision matrix Θ toward a diagonal matrix when $n > p + 4$ is

$$\hat{\Theta}_{OAS} = \hat{\rho}_{OAS} \cdot \text{Diag}(P_n) + (1 - \hat{\rho}_{OAS}) \cdot P_n, \quad (89)$$

$$\hat{\rho}_{OAS} = \min \left\{ \frac{2 \cdot \text{Tr}(\text{Diag}(P_n)^2) + \frac{n-p}{n-p-1} \cdot \text{Tr}(P_n^2) + \frac{n-p-2}{n-p-1} \cdot (\text{Tr} P_n)^2}{\left(\frac{n-p}{n-p-1} + n - p - 4\right) \cdot [\text{Tr}(P_n^2) - \text{Tr}(\text{Diag}(P_n)^2)]}, 1 \right\}. \quad (90)$$

Proof. The proof follows exactly the same lines as in the case of the shrinkage toward identity.

It is omitted. \square

D Optimization algorithm

Global Minimum Variance Portfolio – 150 days

	Mean return	Stand. dev.	Sharpe ratio	Turnover	Herf. ⁻¹	Max. weight	Min. weight	# pos. weight	Short interest	Mean return	Stand. dev.	Sharpe ratio	Turnover	Herf. ⁻¹	Max. weight	Min. weight	# pos. weight	Short interest	
A. Unconstrained optimization																			
Covariance Matrix										Precision Matrix									
<hr/>																			
Basic approaches																			
Sample	23.29	32.44	0.72	27.69	1.62	0.47	-0.47	19.13	-16.52										
			(0.29)		(1.65)	(0.23)	(0.20)	(6.01)	(4.15)										
Identity	15.05	20.88	0.72	0.00	211.00	0.00	0.00	1.00	0.00										
			(0.34)		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)										
Diagonal	11.29	16.79	0.67	0.04	118.38	0.06	0.00	1.00	0.00										
			(0.33)		(52.83)	(0.08)	(0.00)	(0.00)	(0.00)										
Factor models																			
ACP 1F	8.43	9.03	0.93	0.39	19.26	0.03	-0.01	3.74	-1.14	8.43	9.03	0.93	0.39	19.26	0.03	-0.01	3.74	-1.14	
			(0.31)		(7.68)	(0.05)	(0.01)	(5.03)	(1.39)			(0.26)		(7.68)	(0.05)	(0.01)	(5.03)	(1.39)	
ACP 2F	8.69	9.05	0.96	0.47	17.98	0.03	-0.01	3.85	-1.24	8.69	9.05	0.96	0.47	17.98	0.03	-0.01	3.85	-1.24	
			(0.31)		(7.02)	(0.05)	(0.01)	(5.17)	(1.55)			(0.33)		(7.02)	(0.05)	(0.01)	(5.17)	(1.55)	
Market Model	11.09	16.43	0.68	0.21	106.75	0.02	0.00	2.62	-0.01	11.68	17.43	0.67	0.14	114.83	0.06	0	1.01	-0.01	
			(0.31)		(51.78)	(0.03)	(0.00)	(2.34)	(0.02)			(0.30)		(48.16)	(0.07)	(0.00)	(0.03)	(0.03)	
Shrinkage																			
Identity	10.21	8.97	1.14	2.71	12.51	0.07	-0.06	6.07	-3.46	-	-	-	-	-	-	-	-	-	
			(0.28)		(8.06)	(0.02)	(0.02)	(4.15)	(2.03)										
Diagonal	10.24	9.73	1.05	0.16	7.56	0.30	0.00	1.00	0.00	-	-	-	-	-	-	-	-	-	
			(0.28)		(3.09)	(0.14)	(0.00)	(0.00)	(0.00)										
Sparsity																			
	11.3	16.81	0.67	0.09	118.54	0.01	0	2.61	0	10.37	7.81	1.33	0.76	14.94	0.03	0	3.79	-1.18	
			(0.32)		(52.99)	(0.03)	(0.00)	(2.09)	(0.00)			(0.31)		(5.89)	(0.07)	(0.00)	(5.81)	(1.00)	

Global Minimum Variance Portfolio – 150 days (Cont'd)

		b. Constrained optimization																	
		Covariance Matrix									Precision Matrix								
Basic approaches																			
Sample	11.55	13.72	0.84	1.47	51.58	0.04	0	2.61	-										
			(0.28)		(19.69)	(0.01)		(2.23)											
Identity	15.05	20.88	0.72	0.00	211.00	0.00	0	1.00	-										
			(0.34)		(0.00)	(0.00)		(0.00)											
Diagonal	11.29	16.82	0.67	0.04	118.93	0.06	0	1.00	-										
			(0.28)		(52.78)	(0.07)		(0.00)											
Factor models																			
ACP 1F	9.59	9.88	0.97	0.3	8.75	0.04	0	2.61	-	9.59	9.88	0.97	0.3	8.75	0.04	0	2.61	-	
			(0.28)		(3.89)	(0.09)		(7.69)				(0.32)		(3.89)	(0.09)		(7.69)		
ACP 2F	9.06	9.82	0.92	0.32	8.5	0.04	0	2.61	-	9.06	9.82	0.92	0.32	8.5	0.04	0	2.61	-	
			(0.34)		(3.67)	(0.09)		(7.83)				(0.31)		(3.67)	(0.09)		(7.83)		
Market Model	11.18	16.47	0.68	0.21	107.35	0.02	0	2.61	-	11.61	17.45	0.67	0.12	117.94	0.06	0	1.00	-	
			(0.27)		(51.84)	(0.03)		(2.33)				(0.29)		(48.21)	(0.07)		(0.00)		
Shrinkage																			
Identity	11.3	9.82	1.15	0.42	12.51	0.05	0	2.61	-	-	-	-	-	-	-	-	-	-	-
			(0.30)		(5.00)	(0.07)		(6.36)											
Diagonal	10.24	9.73	1.05	0.16	7.56	0.30	0	1.00	-	-	-	-	-	-	-	-	-	-	-
			(0.28)		(3.09)	(0.14)		(0.00)											
Sparsity																			
	11.5	13.99	0.82	0.39	67.14	0.02	0	2.61	-	8.91	10.03	0.89	0.41	11.85	0.05	0	2.61	-	
			(0.31)		(30.35)	(0.04)		(3.10)				(0.29)		(9.97)	(0.12)		(7.01)		

Table 1: The table presents the result of the empirical study for a rolling window of 150 days. Panel A reports the results when short-sells are allowed (Unconstrained optimization) while Panel B reports the results in the absence of short-sells (Constrained optimization). For each panel, the left sub-panel exposes the out-of-sample characteristics of the GMVP when the input parameter is the estimated covariance matrix while the right sub-panel pertains to the results obtained with the estimated precision matrix. Each sub-panel report the out-of-sample mean return, standard deviation, Sharpe ratio, turnover rate, average inverse Herfindal index, largest and smallest weights, number of positive weights and short interests. Figures within parenthesis represents the estimation error. Figures in bold highlight the maximum value of the column.

Global Minimum Variance Portfolio – 200 days

	Mean return	Stand. dev.	Sharpe ratio	Turnover	Herf. ⁻¹	Max. weight	Min. weight	# pos. weight	Short interest	Mean return	Stand. dev.	Sharpe ratio	Turnover	Herf. ⁻¹	Max. weight	Min. weight	# pos. weight	Short interest	
A. Unconstrained optimization																			
	Covariance Matrix									Precision Matrix									
<hr/>																			
Basic approaches																			
Sample	10.00	43.48	0.23	23.57	0.50	0.48	-0.44	11.89	-10.89										
			(0.27)		(0.50)	(0.28)	(0.27)	(6.55)	(6.55)										
Identity	15.05	20.88	0.72	0.00	211.00	0.00	0.00	1.00	0.00										
			(0.34)		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)										
Diagonal	12.45	16.93	0.74	0.03	121.42	0.06	0.00	1.00	0.00										
			(0.31)		(52.39)	(0.07)	(0.00)	(0.00)	(0.00)										
Factor models																			
PCA 1F	9.32	9.16	1.02	0.12	19.07	0.13	-0.02	1.45	-0.45	9.32	9.16	1.02	0.12	19.07	0.13	-0.02	1.45	-0.45	
			(0.28)		(7.14)	(0.10)	(0.01)	(0.20)	(0.20)			(0.28)		(7.14)	(0.10)	(0.01)	(0.20)	(0.20)	
PCA 2F	9.48	9.24	1.03	0.13	18.02	0.13	-0.02	1.49	-0.49	9.48	9.24	1.03	0.13	18.02	0.13	-0.02	1.49	-0.49	
			(0.28)		(6.55)	(0.10)	(0.01)	(0.22)	(0.22)			(0.28)		(6.55)	(0.10)	(0.01)	(0.22)	(0.22)	
Market Model	12.34	16.60	0.74	0.06	111.32	0.07	0.00	1.00	0.00	12.65	17.34	0.73	0.08	123.03	0.05	0.00	1.00	0.00	
			(0.30)		(52.02)	(0.08)	(0.00)	(0.01)	(0.01)			(0.31)		(49.74)	(0.06)	(0.00)	(0.00)	(0.00)	
Shrinkage																			
Identity	9.99	9.04	1.11	1.05	10.07	0.1	-0.06	2.52	-1.52	-	-	-	-	-	-	-	-	-	-
			(0.35)		(6.73)	(0.03)	(0.03)	(0.67)	(0.67)			-	-	-	-	-	-	-	-
Diagonal	9.14	8.48	1.08	0.97	5.57	0.23	-0.05	2.42	-1.42	-	-	-	-	-	-	-	-	-	-
			(0.32)		(1.91)	(0.12)	(0.03)	(0.70)	(0.70)			-	-	-	-	-	-	-	-
Sparsity																			
Sparsity	12.47	16.95	0.74	0.03	121.6	0.06	0.00	1.00	0.00	12.24	11.63	1.05	0.08	48.48	0.1	-0.01	1.08	-0.08	
			(0.30)		(52.56)	(0.07)	(0.00)	(0.00)	(0.00)			(0.28)		(19.49)	(0.08)	(0.00)	(0.04)	(0.04)	

Global Minimum Variance Portfolio – 200 days (Cont'd)

		b. Constrained optimization																	
		Covariance Matrix									Precision Matrix								
Basic approaches																			
Sample	13.96	14.87	0.94	0.78	80.42	0.07	0	1.02	-										
			(0.32)		(69.28)	(0.05)		(0.04)											
Identity	15.05	20.88	0.72	0.00	211.00	0.00	0	1.00	-										
			(0.34)		(0.00)	(0.00)		(0.00)											
Diagonal	12.48	16.96	0.74	0.03	121.97	0.06	0	1	-										
			(0.29)		(52.35)	(0.07)		(0.00)											
Factor models																			
PCA 1F	10.25	9.93	1.03	0.09	8.57	0.27	0	1.00	-	10.25	9.93	1.03	0.09	8.57	0.27	0	1.00	-	
			(0.34)		(3.46)	(0.12)		(0.00)				(0.34)		(3.46)	(0.12)		(0.00)		
PCA 2F	9.84	9.89	1.00	0.09	8.38	0.27	0	1.00	-	9.84	9.89	1.00	0.09	8.38	0.27	0	1.00	-	
			(0.33)		(3.25)	(0.12)		(0.00)				(0.33)		(3.25)	(0.12)		(0.00)		
Market Model	12.41	16.65	0.75	0.06	111.87	0.06	0	1.00	-	12.69	17.36	0.73	0.08	123.98	0.05	0	1.00	-	
			(0.30)		(52.06)	(0.08)		(0.00)				(0.29)		(49.78)	(0.06)	(0.00)	(0.00)		
Shrinkage																			
Identity	11.62	9.86	1.18	0.13	11.13	0.2	0	1.00	-	-	-	-	-	-	-	-	-	-	-
			(0.32)		(3.60)	(0.07)	(0.00)	(0.00)											
Diagonal	10.68	9.84	1.09	0.13	7.5	0.3	0	1.00	-	-	-	-	-	-	-	-	-	-	-
			(0.36)		(2.87)	(0.13)	(0.00)	(0.00)											
Sparsity																			
	12.51	14.99	0.83	0.06	93.94	0.07	0	1.00	-	12.38	12.34	1.00	0.06	48.36	0.10	0	1.00	-	
			(0.31)		(43.26)	(0.07)	(0.00)	(0.00)				(0.30)		(20.31)	(0.08)	(0.00)	(0.00)		

Table 2: The table presents the result of the empirical study for a rolling window of 200 days. Panel A reports the results when short-sells are allowed (Unconstrained optimization) while Panel B reports the results in the absence of short-sells (Constrained optimization). For each panel, the left sub-panel exposes the out-of-sample characteristics of the GMVP when the input parameter is the estimated covariance matrix while the right sub-panel pertains to the results obtained with the estimated precision matrix. Each sub-panel report the out-of-sample mean return, standard deviation, Sharpe ratio, turnover rate, average inverse Herfindal index, largest and smallest weights, number of positive weights and short interests. Figures within parenthesis represents the estimation error. Figures in bold highlight the maximum value of the column.

Global Minimum Variance Portfolio – 350 days

	Mean return	Stand. dev.	Sharpe ratio	Turnover	Herf. ⁻¹	Max. weight	Min. weight	# pos. weight	Short interest	Mean return	Stand. dev.	Sharpe ratio	Turnover	Herf. ⁻¹	Max. weight	Min. weight	# pos. weight	Short interest	
A. Unconstrained optimization																			
Covariance Matrix										Precision Matrix									
<hr/>																			
Basic approaches																			
Sample	8.37	10.42	0.80	2.88	3.78	0.09	-0.07	7.41	-4.99										
			(0.33)		(1.20)	(0.06)	(0.03)	(6.99)	(3.42)										
Identity	15.05	20.88	0.72	0.00	211.00	0.00	0.00	1.00	0.00										
			(0.34)		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)										
Diagonal	12.45	17.44	0.71	0.02	128.81	0.05	0.00	1.00	0.00										
			(0.34)		(49.44)	(0.05)	(0.00)	(0.00)	(0.00)										
Factor models																			
ACP 1F	9.87	9.93	0.99	0.18	19.09	0.02	-0.01	3.6	-1.18	9.87	9.93	0.99	0.18	19.09	0.02	-0.01	3.60	-1.18	
			(0.32)		(5.45)	(0.03)	(0.01)	(4.78)	(1.56)			(0.32)		(5.45)	(0.03)	(0.01)	(4.78)	(1.56)	
ACP 2F	10.26	10.01	1.02	0.21	18.26	0.02	-0.01	3.68	-1.26	10.26	10.01	1.02	0.21	18.26	0.02	-0.01	3.68	-1.26	
			(0.34)		(5.28)	(0.03)	(0.01)	(4.90)	(1.76)			(0.34)		(5.28)	(0.03)	(0.01)	(4.90)	(1.76)	
Market Model	12.18	17.17	0.71	0.08	119.84	0.01	0.00	2.42	0.00	12.74	17.73	0.72	0.04	134.11	0.04	0.00	1.00	0.00	
			(0.30)		(50.91)	(0.02)	(0.00)	(1.92)	(0.01)			(0.31)		(47.00)	(0.04)	(0.00)	(0.00)	(0.00)	
Shrinkage																			
Identity	10.24	9.18	1.11	1.77	7.98	0.06	-0.05	6.10	-3.69	14.63	20.68	0.71	0.04	193.27	0.01	0.00	2.43	-0.01	
			(0.34)		(4.39)	(0.03)	(0.02)	(5.69)	(2.69)			(0.31)		(46.84)	(0.02)	(0.00)	(0.46)	(0.04)	
Diagonal	9.65	8.81	1.09	0.68	5.54	0.21	-0.06	2.43	-1.43	12.00	19.28	0.62	0.04	117.56	0.04	0.00	1.00	0.00	
			(0.34)		(1.26)	(0.08)	(0.02)	(0.64)	(0.64)			(0.29)		(32.99)	(0.03)	(0.00)	(0.00)	(0.00)	
Sparsity																			
										11.2	10.1	1.11	0.21	30.11	0.02	0.00	2.9	-0.48	
												(0.34)		(11.53)	(0.03)	(0.00)	(4.25)	(0.57)	

Global Minimum Variance Portfolio – 350 days (Cont'd)

		b. Constrained optimization																	
		Covariance Matrix									Precision Matrix								
Basic approaches																			
Sample	8.82	10.26	0.86	0.20	7.56	0.03	0	2.42	-										
			(0.32)		(2.47)	(0.07)		(8.11)											
Identity	15.05	20.88	0.72	0.00	211.00	0.00	0	1.00	-										
			(0.34)		(0.00)	(0.00)		(0.00)											
Diagonal	12.48	17.46	0.71	0.02	129.34	0.05	0	1.00	-										
			(0.33)		(49.34)	(0.05)		(0.00)											
Factor models																			
ACP 1F	8.90	10.43	0.85	0.13	8.55	0.03	0	2.42	-	8.90	10.43	0.85	0.13	8.55	0.03	0	2.42	-	
			(0.32)		(2.43)	(0.06)		(7.53)				(0.33)		(2.43)	(0.06)		(7.53)		
ACP 2F	8.41	10.41	0.81	0.14	8.42	0.03	0	2.42	-	8.41	10.41	0.81	0.14	8.42	0.03	0	2.42	-	
			(0.33)		(2.35)	(0.06)		(7.62)				(0.33)		(2.35)	(0.06)		(7.62)		
Market Model	12.20	17.20	0.71	0.08	120.32	0.01	0	2.42	-	12.77	17.74	0.72	0.04	134.62	0.04	0	1	-	
			(0.30)		(50.85)	(0.02)		(1.90)				(0.29)		(46.86)	(0.04)		(0.00)		
Shrinkage																			
Identity	9.7	10.25	0.95	0.19	9.60	0.03	0	2.42	-	14.66	20.71	0.71	0.03	194.79	0.01	0	2.42	-	
			(0.31)		(1.96)	(0.05)		(7.20)				(0.33)		(42.84)	(0.01)		(0.40)		
Diagonal	8.85	10.25	0.86	0.08	7.77	0.27	0	1.00	-	12.04	19.3	0.62	0.04	118.74	0.04	0	1	-	
			(0.34)		(2.53)	(0.10)		(0.00)				(0.34)		(2.53)	(0.10)		(0.00)		
Sparsity																			
										-	11.02	11.64	0.95	0.14	30.19	0.02	0	2.42	-
													(0.29)	(13.20)	(0.04)		(4.51)		

Table 3: The table presents the result of the empirical study for a rolling window of 350 days. Panel A reports the results when short-sells are allowed (Unconstrained optimization) while Panel B reports the results in the absence of short-sells (Constrained optimization). For each panel, the left sub-panel exposes the out-of-sample characteristics of the GMVP when the input parameter is the estimated covariance matrix while the right sub-panel pertains to the results obtained with the estimated precision matrix. Each sub-panel report the out-of-sample mean return, standard deviation, Sharpe ratio, turnover rate, average inverse Herfindal index, largest and smallest weights, number of positive weights and short interests. Figures within parenthesis represents the estimation error. Figures in bold highlight the maximum value of the column.